
OCCASIONAL PAPERS



How to Regulate (and Not Regulate) Social Media

By Jack M. Balkin



KNIGHT
FIRST AMENDMENT
INSTITUTE

at Columbia University

The Knight Institute's Occasional Papers series aims to bring to a broad audience thoughtful, provocative work from scholars and experts who usually write for more specialized audiences. The open-ended series features papers that address urgent questions at the intersection of speech, privacy, and technology.

The series is conceptualized by Knight Institute staff, including Jameel Jaffer, Executive Director; Katy Glenn Bass, Research Director; Alex Abdo, Litigation Director; and Larry Siems, Chief of Staff. The series is edited by Glenn Bass with support from Lorraine Kenny, Communications Director; Sarah Guinee, Research Fellow; and Madeline Wood, Communications and Research Coordinator.

The full series is available at knightcolumbia.org/research/

INTRODUCTION

TO UNDERSTAND HOW to regulate social media, you have to understand why you want to regulate it. I will say something about specific regulatory proposals in the last part of this essay. But I want to spend most of my time discussing the why as much as the how.

Here is the central idea: Social media companies are key institutions in the 21st century digital public sphere. A public sphere doesn't work properly without trusted and trustworthy institutions guided by professional and public-regarding norms. The goal of regulating social media is to create incentives for social media companies to be responsible and trustworthy institutions that will help foster a healthy and vibrant digital public sphere.

What is the public sphere? For purposes of this essay, we can say that the public sphere is the space in which people express opinions and exchange views that judge what is going on in society. Put another way, the public sphere is a set of social practices and institutions in which ideas and opinions circulate. The public sphere is obviously crucial to democracy. But most people's opinions aren't about government policy. They are about sports, culture, fashion, gossip, commerce, and so on.

A public sphere is more than just people sitting around talking. It is shaped and governed, and made functional or dysfunctional, rich or poor, by *institutions*. Most of the institutions that constitute the public sphere are private. They sit between the public and the government. There are lots of examples in the pre-digital world: print and broadcast media, book clubs, spaces for assembly and conversation, sports stadiums, theaters, schools, universities, churches, libraries, archives, museums, and so on.

A digital public sphere is a public sphere that is dominated by digital media and digital technologies. Digital media become the key institutions that either maintain or undermine the health of the public sphere.

THREE KINDS OF DIGITAL SERVICES

BEFORE DISCUSSING HOW we should regulate social media, I want to distinguish social media from two other parts of the infrastructure of digital communication.¹ These are:

(1) Basic internet services, such as the Domain Name System (DNS), broadband companies, and caching services.

(2) Payment systems, such as MasterCard, Visa, and PayPal.

For basic internet services the regulatory answer is pretty simple: Non-discrimination. Let the bits flow freely and efficiently. Don't try to engage in content regulation at this level. Government should enforce non-discrimination as a matter of policy. Although the question is contested (for example, in the policy debates over network neutrality rules), I believe that enforcing nondiscrimination rules at this level of the internet presents no significant First Amendment problems.

We should treat payment systems, and caching and defense systems, like public accommodations, with this caveat: They can refuse to do business if a customer uses their business for illegal activities.

Governments and civil society groups often want to use basic internet services and payment systems to go after propagandists, conspiracy mongers, and racist speakers. I think this is a mistake. These businesses are not well designed for content moderation and their decisions will be arbitrary and ad hoc.

I believe that content regulation should occur higher up the stack, to borrow a familiar computer science metaphor.

Instead, these businesses should concern themselves only with the legality or illegality of transactions. Government should require nondiscrimination—otherwise the public and politicians will place irresistible pressure on basic internet services and payment systems to engage in content moderation, which is not their job.

Government requirements of nondiscrimination/public accommodation have this advantage: when civil society groups and politicians demand that these businesses engage in content moderation, or argue that businesses are complicit in the politics of the customers they serve, the businesses can respond that they have no choice because the law requires them not to discriminate against customers who are not engaged in illegal activity.

Instead, content moderation should occur in social media and search engines. In fact, for these services, content regulation is inevitable. Since it is inevitable, that's where you should do it.

THE PUBLIC FUNCTION OF SOCIAL MEDIA

NOW LET'S ASK: what is social media's public function? What tasks should it perform in the digital public sphere?

This is a normative and interpretive question. So too is the related question of what it means for the public sphere to be well functioning, "healthy," or "vibrant." We must decide what makes the digital public sphere function well or badly. Because social media are so new, we have very little history to work with. So we have to make analogies to the longer history of media and democracy. But in doing so, we also have to reckon with the fact that earlier versions of the public sphere may not have functioned well.

I mentioned previously that the public sphere created by social media in the 21st century is a successor to the public sphere created by print and broadcast media in the 20th century. Twentieth-century media helped produce a particular kind of public sphere, different than today's, because broadcast and print media played a different role than social media do today. These companies—or their contractual partners—produced most of the content that

they published or broadcast. Twentieth-century print and broadcast media were not participatory media; the vast majority of people were audiences for the media rather than creators who had access to and used the media to communicate with others.

The 21st century model, by contrast, involves crowdsourcing and facilitating end user content. Social media host content made by large numbers of people, who are both creators and audiences for the content they produce.

If that's so, what are social media's central functions in the public sphere? What is social media's appropriate role? I argue that social media have three central functions:

First, social media *facilitate public participation* in art, politics, and culture.

Second, social media *organize public conversation* so people can easily find and communicate with each other.

Third, social media *curate public opinion*, not only through individualized results and feeds, but also through enforcing community standards and terms of service. Social media curate not only by taking down or rearranging content, but also by regulating the speed of propagation and the reach of content.

This last point bears elaboration. During the 20th century, newspapers and television also curated public discourse through the exercise of editorial judgment. They decided what content to commission in the first place and how to edit and convey the content they eventually produced. That meant that the content that circulated in these media was restricted and sanitized for mass audiences. One did not see pornography in The New York Times or advocacy of racial genocide on NBC because these companies had standards and professional norms about what they would publish or broadcast. These standards and norms, in turn, were backed up by legal requirements—for example, against defamation, obscenity, and indecency. Even so, 20th century media companies often limited speech far more than the law required.

Twentieth-century mass media set boundaries on permissible content, and created a certain kind of public conversation based on the expected interests and values of their audiences. Different players in different media and in different parts of society imposed different norms. Book publishers applied their own set of norms, motion picture companies had their own set

of norms, the pornography industry (which encompassed both print and video) had its own norms, and so on. Generally speaking, daily newspapers and broadcast media applied norms of a hypothesized polite society judged appropriate for an imagined audience of average adults and their families. One could get access to more daring content elsewhere, for example in books and magazines, subject always to background legal constraints.

Social media also curate public discourse today. But instead of publishing their own content, they are publishing everyone else's content. Like 20th century mass media, they apply a set of rules and standards about what kinds of content (and conversations) are permissible and impermissible on their sites. They impose a set of civility, safety, and behavioral norms for their imagined audience—different from 20th century newspapers, but nevertheless still quite constrained. Different social media enforce different norms. Like 20th century media, social media may limit speech far more than the law requires them to. Facebook, for example, limits nudity even when it is constitutionally protected.²

Generally speaking, the free speech principle allows the state to impose only a very limited set of civility, safety, and behavioral norms on public discourse, leaving intermediate institutions free to impose stricter norms in accord with their values. This works well if there are many intermediate institutions. The assumption is that in a diverse society with different cultures and subcultures, different communities will create and enforce their own norms, which may be stricter than the state's. I believe that a diversity of different institutions with different norms is a desirable goal for the public sphere in the 21st century too. But I also believe that there is a problem—no matter which century we are talking about—when only one set of norms is enforced or allowed. If private actors are going to impose norms that are stricter than what governments can impose, it is important that there be many different private actors imposing these norms, reflecting different cultures and subcultures, and not just two or three big companies. I will return to this point later on.

Now let me connect the three functions I mentioned—facilitating public participation, organizing public conversation, and curating public opinion—to the goals of a healthy, well-functioning, public sphere. Why are these functions the key indicia of a well-functioning public sphere?

These functions are important because the public sphere is the institutional home of freedom of speech and it helps realize the values of freedom of expression. Free speech values help us understand whether the public sphere is functioning well or badly. If the institutional arrangements work well to facilitate these values, then we say that the public sphere is functioning well, and that it is healthy. But if institutional arrangements hinder these values, we should conclude that the public sphere is not functioning well.

Well, what are these values? There are at least three of them:

First, freedom of speech serves the values of political democracy. It enables *democratic participation* in the formation of public opinion. It helps to ensure (although it does not guarantee) that state power is responsive to the evolution of public opinion. And it helps to ensure (although it does not guarantee) that the public can become informed about issues of public concern. Thus the democratic political values are participation, responsiveness, and an informed public.

Second, freedom of speech helps to produce a *democratic culture*. A democratic culture is a culture in which individuals and groups can freely participate in culture and in the forms of cultural power that shape and affect them.³ Because cultural power is even more pervasive than state power, individuals need to have a way of participating in the construction and development of the cultures that constitute their identities and affect their lives. Freedom of speech allows widespread participation in the forms of meaning making that construct us as individuals. It gives people a chance to talk back to and shape the forms of cultural power that constitute them.

Third, freedom of speech helps promote (although once again, it does not guarantee) *the growth and spread of knowledge*. I use this formula instead of the familiar “marketplace of ideas” because the latter metaphor is misleading. The best way to develop and spread knowledge may not be through competition for acceptance in public opinion. Instead, in modern societies, the development and spread of knowledge depends on a host of disciplines, institutions, and public-regarding professions.

Social media perform their public functions well when they promote these three central values: political democracy, cultural democracy, and the growth and spread of knowledge. More generally, a healthy, well-functioning digital public sphere helps individuals and groups realize these three central

values of free expression. A poorly functioning public sphere, by contrast, undermines political and cultural democracy, and hinders the growth and spread of knowledge.

TRUSTED AND TRUSTWORTHY INTERMEDIATE INSTITUTIONS

HERE'S THE NEXT BIG IDEA: If you want to realize these values, you need more than a simple free speech guarantee like the American First Amendment. You need more than a legal norm that the state doesn't censor. You need more than the formal ability to speak free of government sanction. You need intermediate institutions that can create and foster a public sphere. Without those intermediate institutions, speech practices decay, and the public sphere fails.

A healthy system of free expression requires much more than non-censorship.

First, it requires knowledge institutions and knowledge professionals who produce and disseminate knowledge and opinion. Examples from the 20th century include newspapers and other media organizations, schools, universities, libraries, museums, and archives. Some of these may be run and/or subsidized by the state. But many of them will be privately owned and operated.

Second, you need lots of different institutions, and they can't all be owned or controlled by a small number of people. They have to provide what Justice Hugo Black once called "diverse and antagonistic sources" of information.⁴ This is a famous formula in First Amendment law. But this formula is not just about having lots of different voices that disagree with each other. Rather it's about having lots of different institutions for knowledge production and dissemination.

Third, these institutions have to have professional norms that guide how they produce, organize, and distribute knowledge and opinion.⁵

Fourth, these intermediate institutions and professional groups can successfully do their job only when they are generally trustworthy *and* trusted. When intermediate knowledge producing institutions and professions are

not trusted, the public sphere will begin to fall apart. Why will it begin to fall apart? Because no matter what your theory of free speech might be, realizing the values of free speech depends on the creation, curation, and dissemination of knowledge by intermediate institutions and professions that the public generally trusts. Without these trusted institutions and professions, the practices of free expression become a rhetorical war of all against all. Such a war undermines the values of political democracy, cultural democracy, and the growth and spread of knowledge that free expression is supposed to serve. Protection of the formal right to speak is necessary to a well-functioning public sphere. It is just not sufficient.

In a nutshell, that is the problem we are facing in the 21st century. We have moved into a new kind of public sphere—a digital public sphere—without the connective tissue of the kinds of institutions necessary to safeguard the underlying values of free speech. We lack trusted digital institutions guided by public-regarding professional norms. Even worse, the digital companies that currently exist have contributed to the decline of other trusted institutions and professions for the creation and dissemination of knowledge.

The irony is profound. Never has it been easier to speak, to broadcast to millions. Never has access to the means of communication been so inexpensive and so widely distributed. But without the connective tissue of trusted and trustworthy intermediate institutions guided by professional and public-regarding norms, the values that freedom of speech is designed to serve are increasingly at risk. Antagonistic sources of information do not serve the values of free expression when people don't trust anyone and professional norms dissolve. InfoWars is an antagonistic source of information. Boy, is it antagonistic! But its goal is to destroy trust. Its goal is to get you to trust nobody. It reduces politics to tribalism and cultural participation to warfare. It reverses and undermines the spread and growth of knowledge.

DIVERSE AFFORDANCES, VALUE SYSTEMS, AND INNOVATIONS

TO ACHIEVE A HEALTHY AND VIBRANT PUBLIC SPHERE, we also need many different kinds of social media with many different affordances, and many different ways to participate and make culture. Thus, it is important to have Facebook and YouTube and TikTok and Twitter, and many other kinds of social media applications as well. Moreover, these applications can't be owned or controlled by the same companies.

Diversity of affordances and control is important for three reasons. First, you don't want one set of private norms governing public discourse. Ideally, different social media will set their own community standards and values, even if they overlap to some degree. Second, you want many players because you want continuous innovation. Third, you want many different kinds of social media because different affordances make culture richer and more democratic.

So in addition to "diverse and antagonistic sources of information" we should want "diverse affordances, value systems, and innovations." But, as I said before, "diverse and antagonistic" is not enough. Social media also need to become trusted mediating institutions guided by professional norms. They have to become trusted and trustworthy organizers and curators of public discourse. They aren't now.

One might object: won't network effects doom the goal of a world with many different kinds of social media? Won't people gravitate to one social media application because everyone else they know is already using it?

The answer is no. Many people currently use many different social media applications, not a single one. They belong to several communities and their usage changes over time. There are several reasons for this.

First, social media have different affordances and people use social media for many different purposes. One can be a member of Facebook and still use YouTube or TikTok. If we encourage diversity of affordances, we will also encourage diversity of use.

Second, people may use different social media more or less frequently and move to new social media as they get older or as their tastes and needs change. Younger people may move to different social media than their

parents and grandparents. We have already seen generational migration from MySpace to Facebook and from Facebook to Snapchat and TikTok.

Third, people may link content from one social media site to others; in a tweet, for example, they may link to a YouTube video or a Spotify playlist.

Social media have incentives to allow people to belong to multiple sites because they want people to switch to their application. Moreover, because they want to be useful (and perhaps even indispensable) to end users, they also have incentives to allow links to other parts of the internet, including other social media. Regulation can encourage this kind of openness, too. If we promote innovation among social media companies, with many different kinds of affordances, network effects will not prevent a larger number of players than we currently have.

THE LIMITS OF ECONOMIC INCENTIVES

S O FAR I'VE OFFERED a set of ideals to aim at. I've told you what a healthy digital public sphere would look like. And I've told you what kinds of institutions we might need.

But it's pretty obvious, when we turn to the real world, that social media are not living up to their appropriate roles in the digital public sphere.

Why? Well, social media are driven by market incentives. In fact, sometimes they are so big that they make their own markets. So economic incentives or profit motives are probably more accurate terms than market incentives. The largest social media are less subject to market discipline than other firms; and lack of competition is one important reason why social media don't live up to their social function in the digital public sphere. Yet it is only one part of the problem.

Economic incentives may be necessary for a healthy public sphere, but they will not be sufficient. Here is why: Free expression and the production of knowledge goods produce both positive and negative externalities. That is, they produce benefits and harms that can't be completely captured by ordinary market transactions. The result is that markets—even perfectly functioning competitive markets—will overproduce the harms of free expression and under-produce the goods of free expression. And this is true whether

media goods are financed through advertising, subscription, or pay services.

Whatever your theory of free expression is, market competition won't produce the kind of culture and knowledge necessary for democratic self-government, democratic culture, or the growth and spread of knowledge. Markets will under-produce the kinds of speech and knowledge goods that support political and cultural democracy; they will under-produce the kinds of institutions that will reliably discover and spread knowledge. Conversely, market incentives will overproduce conspiracy theories and speech that undermines democratic institutions. When social media are dominated by a small number of powerful economic actors, their incentives are not much better.

Economic incentives are not the same thing as professional norms and they may come into conflict with and undermine professional norms.

And today, economic incentives for social media companies promote distrust, not trust. They undermine professional norms for the production of knowledge rather than support them.

Then add the fact that all of this takes place on the internet. The internet is just a big machine for destroying professional norms.

It's not surprising that social media have failed at the task I just set out for them. For one thing, they are still very new. Facebook is only a decade and a half old. Google is only 20 years old. They emerged as profit-making technology companies, and only later came to understand themselves as media companies. They were brought to this realization kicking and screaming all the way, through continuous and sustained public pressure.

And yet this is the direction they must travel. Social media companies have to become key institutions for fostering a healthy public sphere. They can't just serve economic incentives. They have to adopt public-regarding professional norms related to the important public function that they serve in the digital public sphere.

By analogy, think about journalism. It also serves a crucial role in the public sphere because it informs the public and sets agendas for public discussion. If the professional norms of journalism are weakened or destroyed and the practice of journalism becomes solely market driven, journalism will make the public sphere worse, not better. It will choose stories and treatments that increase polarization, tribalism, and social distrust, and it

will generate or help spread propaganda and conspiracy theories.

In fact, social media has multiple roles to play in the digital public sphere.

First, social media companies are important players in many different kinds of regulation. Public-private cooperation is necessary for dealing with, among other things, terrorist recruitment, foreign interference in elections, campaign finance violations, and child pornography.

Second, huge digital communities create special problems of personal safety, threats, and abuse. Some countries present special problems of state propaganda and genocidal speech campaigns.

Third, the need for content moderation creates problems of scale. Content moderation that is simultaneously quick, accurate, and at scale is hard to achieve. Accuracy requires increasing the number of moderators (either through hiring or contracting out to other firms) at numbers far greater than most social media companies would like; it also requires treating content moderators much better than they are currently treated by their employers.⁶ In fact, social media companies often rely on complaints by end users, civil society organizations, and government actors to spot violations of their terms of service. Because moderation is costly to do well, social media companies have economic incentives to drag their heels.

MISALIGNED INCENTIVES

ARE THERE INCENTIVES for social media to become trustworthy institutions that protect and foster the digital public sphere? Sadly, not as they are currently constituted.

Social media companies have been slow to solve the problems they create. Social media companies have viewed themselves primarily as technology companies that make money through digital surveillance that enables advertising. Their goal is to get bigger and bigger, and to expand their user base so they can serve more ads and make more money.

The 20th century public sphere was also partly funded through advertising. But its problems were a bit different, because you didn't have modern methods of data collection and behavioral advertising. Also 20th century

media had greater professional and economic incentives to be trustworthy, even if they were hardly perfect and tended to be too passive and apologetic.

Advertising (and therefore data collection and manipulation) are central to the problems that social media creates for the digital public sphere. There are three reasons for this.

First, the attention economy generates perverse effects. It encourages companies to highlight the kind of content that keeps viewers' attention. This content is less likely to be informative or educational, and more likely to be false, demagogic, conspiratorial, and incendiary, and to appeal to emotions such as fear, envy, anger, hatred, and distrust.

Second, Facebook and Google serve both as advertising brokers and as the major market for ads. They are a digital advertising duopoly.

Third, Facebook and Google have dried up revenues for newsgathering organizations, who get an increasingly small amount of ad revenues, or have to take crumbs off the table from Facebook and Google. The internet has created news deserts for local news and increased incentives for consolidation of media organizations into a handful of large companies. Put another way, one side effect of market incentives has been undermining other public sphere institutions—in particular, journalism—and the advertising-based business models that have traditionally sustained journalism.

Economic incentives have driven Facebook and Google to grow ever larger and to buy up as many potential competitors as possible. But a well-functioning digital public sphere should have many social media companies, not just a few, because:

1. you don't want a monoculture of content moderation;
2. having lots of different players in different parts of the world partly eases problems of scale in moderation;
3. many players make it harder for foreign governments to hijack elections;
4. many players may be better for innovation; and
5. many players are harder for governments to co-opt.

To all of these we should add a sixth reason tied to the dangers of surveillance capitalism.⁷ Facebook's and Google's control over digital advertising is

made possible by their ability to collect and aggregate enormous amounts of end user data, more than any other company. The more data Facebook and Google are able to collect, the better their predictive algorithms, the more powerful their ability to nudge and influence end users, and the better their ability to corner the market on digital advertising. That is why it is profitable for Facebook and Google to buy up so many different kinds of companies and applications, each of which collects data in different ways. More data means more power.

If there are many different social media companies, none will have the same dominance and control over the collection and analysis of end user data. None will have the same power to manipulate and influence end users, and none will be able to corner the market for digital advertising. Having more players diffuses and decentralizes power over the collection and control of data, over digital advertising markets, and over end users, who are the objects of surveillance, influence, and manipulation.

PUBLIC PROVISIONING AND PUBLIC UTILITIES

MANY PEOPLE HAVE SUGGESTED public provisioning—state-run social media—as a solution to the problems of social media. Others have suggested turning social media companies into public utilities.

Let's start with public provisioning. Certainly one way to provide public goods that the market will fail to provide adequately is to have government provide it. That's what we do with state universities and what many countries do with public broadcasting.

But unlike state universities and public broadcasters like the BBC, you really don't want governments to provide social media services:

First, if social media companies are treated as state actors and have to abide by existing free speech doctrines—at least in the United States—they will simply not be able to moderate effectively. Facebook's and Twitter's community standards, for example, have many content-based regulations that would be unconstitutional if imposed by government actors. Even if one eliminated some of these rules, the minimum requirements for effective

online moderation would violate the First Amendment.

Second, content moderation does not give speakers final judicial determinations (with full Bill of Rights protections) of whether their speech is protected or unprotected. Therefore content moderation has many of the same problems as administrative prior restraints. The standard remedies for violating community standards and/or terms of service include removing an end user's content and banning the end user from the community. Some of these remedies would probably violate American free speech doctrine, including the rule against prior restraints. If A defames B in a public park, for example, a court could not forbid A from ever speaking in the park again.⁸

Third, and relatedly, many people are concerned about the propagation of false and misleading political advertisements and political propaganda on social media. They want social media companies to take down this speech or prevent it from being used in targeted political messages and ads. But if that is your concern, the last thing you would want to do is make social media state actors, because state actors are severely constrained in how they can sanction political speech, even false political speech. And, once again, even when state actors may sanction political speech, they must first afford the speaker the full panoply of Bill of Rights protections and a final individualized judicial determination before they can act. These requirements are simply inconsistent with the speed and scale of social media content moderation.

Fourth, if you think that surveillance capitalism is bad, there are even more serious problems of government surveillance and data manipulation when governments run your social media company.

Fifth, and relatedly, governments running social media services would create enormous risks of facilitating government propaganda and the use of end user data to engage in targeted influence campaigns.

Sixth, and finally, governments may not be particularly good at innovation. And they will not be very good at facilitating a diverse set of affordances, values, and innovations.

Another approach is to turn social media and search engines into privately owned public utilities.⁹

It is not clear that social media fit the traditional model of public utilities very well. The classic examples of public utilities are companies that provide water, telephone services, and electrical power. The standard reasons for

making a company a public utility are to control price, to secure universal access, and to assure the quality of continuous service. But with social media, the price is free, access is universal, and continuous service is almost always provided—in part because companies want as much of end users’ attentions as they can get. If the real goal of treating social media as public utilities is to prevent discrimination in content moderation, then one faces the same problems as state-run social media.

Probably the best justification for a public utility model is to fundamentally change the business model of social media companies. Once converted into public utilities, social media companies would give up advertising altogether and simply provide access and content moderation services in return for a fixed monthly subscription fee. (They might still be allowed to run ads, but the ads could not be targeted.) This arrangement would have to be combined with strict limits on collection, collation, and sale of end user data. That is because the mere fact that subscription services don’t serve you ads doesn’t mean that they respect your privacy or are not attempting to manipulate you; they might continue to collect end user data and sell it to other companies or use it for other purposes.

It may well be a good idea to have some subscription-based social media services in a larger mix of social companies that rely on advertising. These social media companies would be a sort of “public option” that people who want extra privacy protections could use as an alternative to free services. But the public utility model is not a general solution to the problems of the digital public sphere. Converting all large social media companies into public utilities does not solve the problems I mentioned above, because it does not provide diverse affordances, value systems, and innovations. Quite the contrary: converting social media companies into public utilities appears to concede that there will only be—and perhaps should only be—a relative handful of social media companies. The more important focus of regulation, therefore, should be on antitrust, privacy, and consumer protection regulation, as I explain below.

TO CHANGE INCENTIVES, CHANGE BUSINESS MODELS

I EXPECT THAT most social media companies will continue to be privately owned and operated, and they will still rely on advertising models. If so, how is it possible to push privately owned social media companies to fulfill their proper social function?

We are slowly inching toward this approach. Social media companies already assert in their public relations materials that they have obligations to the public. They state that they understand that their businesses depend on public trust. They acknowledge that it is their goal to protect end user autonomy, enhance democracy, and facilitate free speech. They make similar claims in their terms of service and community standards. Whether social media companies actually live up to these claims is more complicated. That is because social media companies are not really willing to give up control of their “crown jewels”: business models based on data collection, behavioral advertising, and other aspects of surveillance capitalism.

Public pressure and media coverage of social media companies can push them, at the margins, to behave as more responsible curators of public discourse. (I should also say that people push social media to be irresponsible and arbitrary as well.) This sort of pressure is important because social media companies don’t want to lose their base of end users. But regulation is also necessary.

Facebook’s Oversight Board for Content Decisions is yet another strategy to generate public trust by attempting to establish a kind of legitimacy in its content moderation decisions. Facebook hopes to use the model of a supreme court—complete with cases, judges and decisions—to establish that Facebook is a trustworthy, public-regarding institution.

I have no objection to the Board in theory. We should encourage every reform that gives social media companies incentives to act in a public-regarding fashion. As currently imagined, however, the Oversight Board won’t be able to do very much. It will consider only a tiny fraction of the content moderated on Facebook in a given year. More importantly, it will have no jurisdiction over Facebook’s crown jewels: the company’s system for brokering advertisements, its behavioral manipulation of end users, and its

practices of data surveillance, collection, and use. For this reason, there is a very real danger that the Oversight Board will prove to be little more than a digital Potemkin Village—a prominent display of public-spiritedness that does nothing to address the larger, deeper problems with social media.

The logic of social media business models will tend to overcome any public statements of ideals, good will, and promises of good behavior. This has happened over and over again. Facebook’s history as a company has been a cycle of engaging in bad behavior, getting caught, apologizing profusely and promising to mend its ways, followed by the company engaging in slightly different bad behavior, offering new apologies and promises of reform, and so on.¹⁰ Facebook will keep misbehaving and it will keep apologizing, not because it is incompetent or clumsy, but because of a fundamental misalignment of incentives between its goals and the public’s needs, and because it has an inherent conflict of interest with its end users and, indeed, with democracy itself.

Social media companies will behave badly as long as their business models cause them to. Profit-making firms like Facebook will normally seek to externalize as many costs of their activities as possible onto others, so that the costs will be borne by society. Their business models don’t care about your democracy.

How do you make social media companies responsible participants in the digital public sphere? First, you must give them incentives to adopt professional and public-regarding norms. Second, you must make them internalize some of the costs they impose on the world around them.

There are no complete, perfect solutions. But we can make progress in incremental steps.

Before I discuss reform strategies, however, there is an important threshold question: Can the U.S. do this on its own? After all, anything we do in the U.S. will be affected by what other countries and the EU do. Today, the EU, China, and the U.S. collectively shape much of internet policy. They are the three Empires of the internet, and other countries mostly operate in their wake. Each Empire has different values and incentives, and each operates on the internet in a different way. I could write an entire essay just on these problems.

MODELS FOR REGULATION

IN THE REMAINDER of this essay, however, I will assume that the U.S. government—and the 50 state governments—can do something on their own. If so, what kinds of regulation should the U.S. consider?

First, don't rush to impose direct regulation on social media moderation practices. Requiring "neutrality" in content moderation is a non-starter. As I explained earlier, neutrality should apply lower down in the stack—to basic internet services—and to payment systems. One of the ironies of the current policy debate is that the very politicians who call for neutrality in content moderation have been most opposed to requiring neutrality where it is most needed—in basic internet services such as broadband.

Social media platforms must engage in content moderation. They may do it badly or well, but they will have to do it nevertheless.¹¹ Accordingly, governments should respect social media's role as curators and editors of public discourse. Respecting that role means that social media should have editorial rights, which are a subset of free speech rights.

The goal of regulation is not to achieve an illusory neutrality in social media content moderation. Rather the goal is to shape the organization and incentives of the industry to better achieve public ends.

First, the goal should be to increase the number of players, so there can be many different companies, communities, affordances, and editorial policies.

Second, the goal should be to give social media companies incentives to professionalize and take responsibility for the health of the public sphere.

We can regulate social media using three policy levers.

1. Antitrust and competition law
2. Privacy and consumer protection law
3. Balancing intermediary liability with intermediary immunity.

Properly structured, none of these policy levers violate free speech values or the First Amendment.

Whatever we do, it is important to keep regulatory burdens manageable. If you make the regulatory burdens too great, you can create barriers

to entry for new social media firms, which defeats the regulatory purpose of achieving a wide range of social media companies with different rules, affordances, and innovations.

Let me talk about antitrust, privacy, and intermediary liability in turn. The discussion that follows will be very broad brush and pitched at a high level of abstraction. I emphasize at the outset that you need all three of these policy levers to succeed. You can't rely on just one. For example, if you don't use antitrust law and competition law, you will have to regulate more heavily in other ways.

Moreover, there are some kinds of problems that privacy law can't fix and for which antitrust law is required; conversely, there are problems that antitrust law can't fix that require privacy and consumer protection law. For example, even if you create many different Facebooks and Googles, each will still be practicing their own forms of surveillance capitalism. You will still need privacy and consumer protection regulations to keep these smaller companies from manipulating and/or abusing the trust of end users.

Antitrust and Competition Law

In competition policy, the goal is not simply separating existing social media services owned by a single company, for example, separating Facebook from Instagram and WhatsApp or YouTube from Google. Rather, there are three interlocking goals.

First, competition policy should aim at producing many smaller companies, with different applications, communities, and norms. You might think of this as a sort of social media federalism.

Second, competition policy should seek to prevent new startups from being bought up early. This helps innovation. It prevents large companies from buying up potential competitors and killing off innovations that are not consistent with their current business models.

Third, competition policy should seek to separate different functions that are currently housed in the same company. This goal of separation of functions is different from a focus on questions of company size and market share.

For example, Facebook and Google are not just social media companies, they are also advertising agencies. They are both Don Draper and NBC. They

match companies who want to advertise with audiences they create, and then they serve ads to end users on their social media feeds and applications.

Hence, competition policy might seek to separate control over advertising brokering from the tasks of serving ads, delivering content, and moderating content. Each of these functions is currently housed in a single company, but some of these tasks could be performed by different companies, each in a separate market.

Conversely, we might want to relax antitrust rules to allow media organizations to collectively bargain with social media companies for advertising rates and advertising placements.

I use the term competition law in addition to antitrust law for a reason. In the United States, at least, antitrust law generally refers to the judicial elaboration of existing antitrust statutes. But in dealing with the problems that social media create for the public sphere, we should not limit ourselves simply to elaborating the current judge-made doctrines of antitrust law, which focus on consumer welfare. Even if we expand the focus of antitrust law to the exercise of economic power more generally, competition law has other purposes besides fostering economic competition, economic efficiency, and innovation. In telecommunications law, for example, media concentration rules have always been concerned with the goal of protecting democracy, and with the goal of producing an informed public with access to many different sources of culture and information. Existing judge-made doctrines of antitrust law might not be the best way to achieve these ends, because they are not centrally concerned with these ends. We might need new statutes and regulatory schemes that focus on the special problems that digital companies pose for democracy.

Privacy and Consumer Protection

I have written a great deal about how we might rethink privacy in the digital age and I won't repeat all of my arguments here.¹² My central argument is that we should use a fiduciary model to regulate digital companies, including both social media companies and basic internet services that collect end user data. A fiduciary model treats digital companies that collect and use data as information fiduciaries toward the people whose data they collect and use.

Information fiduciaries have three basic duties towards the people

whose data they collect: a duty of care, a duty of confidentiality, and a duty of loyalty. The fiduciary model is not designed to directly alter content moderation practices, although it may have indirect effects on them. Rather, the goal of a fiduciary model is to change how digital companies, including social media companies, think about their end users and their obligations to their end users. Currently, end users are treated as a product or a commodity sold to advertisers. The point of the fiduciary model is to make companies stop viewing their end users as objects of manipulation—as a pair of eyeballs attached to a wallet, captured, pushed, and prodded for purposes of profit.

This has important consequences for how companies engage in surveillance capitalism. If we impose fiduciary obligations, even modest ones, business models will have to change, and companies will have to take into account the effects of their practices on the people who use their services.

The fiduciary model is designed to be flexible. It can be imposed by statute, through administrative regulation, or through judicial doctrines. Fiduciary obligations are one important element of digital privacy and consumer protection but they are not sufficient in and of themselves. Moreover, fiduciary obligations must work hand in hand with competition law, because each can achieve things that the other cannot.

Intermediary Liability

One of the central debates in internet law is whether and how much intermediary liability states should impose, and conversely, whether states should grant some form of intermediary immunity. In general, I believe that intermediary immunity is a good idea, and some (but not complete) intermediary immunity is actually required by the free speech principle.

Because the current broad scope of intermediary immunity is not required by the First Amendment or the free speech principle more generally, governments should use the offer of intermediary immunity as a lever to get social media companies to engage in public-regarding behavior. In particular, one should use intermediary immunity as a lever to get social media companies to accept fiduciary obligations toward their end users.

Governments might also condition intermediary immunity on accepting obligations of due process and transparency. Social media companies currently have insufficient incentives to invest in moderation services and

to ensure that their moderators are treated properly. In some cases, governments might be able to regulate the provision of moderation services through employment and labor law (although there are a few free speech problems with media-specific regulations that I can't get into here). But governments should also create incentives for platforms to invest in increasing the number of moderators they employ as well as providing more due process for end users. They should also require companies to hire independent inspectors or ombudsmen to audit the company's moderation practices on a regular basis.¹³ In short, I don't want to scrap intermediary immunity. I want to use it to create incentives for good behavior.

Although the general rule should be intermediary immunity, governments may partially withdraw intermediary immunity and establish distributor liability in certain situations. Distributor liability means that companies are immune from liability until they receive notice that content is unlawful. Then they have to take down the content within a particular period of time or else they are potentially vulnerable to liability (although they may have defenses under substantive law).

First, governments might employ distributor liability for certain kinds of privacy violations; the most obvious example is non-consensual pornography, sometimes called "revenge porn."

Second, governments might establish distributor liability for paid advertisements. The basic problem of intermediary liability—and the reason why intermediary immunity is a good thing—is the problem of collateral censorship. Because companies can't supervise everything that is being posted on their sites, once they face the prospect of intermediary liability they will take down too much content, because it is not their speech and they have insufficient incentives to protect it. This logic does not apply in the same way, however, for paid advertisements. Companies actively solicit paid advertisements—indeed, this is how social media companies make most of their money. As a result, even with distributor liability, companies still have incentives to continue to run ads. These incentives lessen (although they do not completely eliminate) the problems of collateral censorship. Note that the rule of distributor liability is still more generous than the rule of publisher liability that currently applies to print media advertisements.

This approach does not require us to distinguish between commercial

advertisements and political advertisements. Nor does it require us to distinguish between issue ads and ads that mention a particular candidate. The on/off switch is simply whether the company accepts advertising. This rule leaves matters up to the company to decide how best to handle advertising, which is, after all, the core of its business. Twitter has recently announced that it will no longer accept political advertisements.¹⁴ Facebook's policies are more complicated and currently in flux. Facebook does take down paid political ads that lie about polling times and places. But it will not take down other false political ads, even when Facebook knows that they are false.¹⁵

Facebook's case is instructive for how to think about the problem. Facebook argues that it does not want to be the arbiter of public discourse. In fact, it already is the arbiter of public discourse worldwide; moreover, as I've argued above, its proper function as a social media company is to serve as a curator of public discourse. Facebook well understands this: it takes down lies about election dates and polling places; and it bans abusive and dehumanizing speech that would otherwise be protected under the First Amendment. It is true that policing political advertisements poses genuine problems of scale: Facebook would have to take down ads not only for federal elections in the U.S., but for every state and local government election, and for every election around the world. However, Facebook already invests in moderating a far larger class of non-advertising speech around the world. So it would have to show why moderating the far smaller class of advertisements—which are marked and inserted into end users' feeds as advertisements—is significantly more difficult.

The real reasons why Facebook has decided not to take down false political ads are somewhat different, and they better explain Facebook's incentives to host political ads. That is important because, as noted above, distributor liability is less troublesome from a free speech perspective when companies have independent incentives to protect certain speech and prevent it from being removed.

First, Facebook probably resists taking down false political advertisements because it makes money from these ads, perhaps more money than it lets on. It is, after all, an advertising company, and unless the law imposes costs for running advertisements, each advertisement adds to its bottom line. But political advertising is only a small fraction of its business, and so

ad revenue is probably not the central motivating factor behind Facebook’s policies. A second and more important reason is that Facebook does not want to anger the politicians who place political ads, and who might be motivated to regulate or break up the company. Regulation or breakup might truly threaten Facebook’s revenues.

Third, Facebook is in the influence business. Serving political ads keeps Facebook connected to important politicians and political actors around the world and thereby increases the company’s power and political influence. That is one reason—although certainly not the only reason—why Facebook treats important political figures differently than ordinary individuals, and keeps up postings that would otherwise violate its community standards or terms of service if made by ordinary individuals.¹⁶ Facebook believes that people want to know what these important figures think; but more importantly, it wants to be the conduit for people to hear what these important people have to say. It also wants to stay on the good side of powerful people who might someday threaten its business. Because Facebook has incentives to solicit, attract, and keep up political advertisements, including knowingly false political advertisements, imposing distributor liability for all advertisements will give Facebook better incentives than it currently has.

CONCLUSION

THE THESIS OF THIS ESSAY is that you shouldn’t regulate social media unless you understand why you want to regulate it.

We should regulate social media because we care about the digital public sphere. Social media have already constructed a digital public sphere in which they are the most important players. Our goal should be to make that digital public sphere vibrant and healthy, so that it furthers the goals of the free speech principle—political democracy, cultural democracy, and the growth and spread of knowledge. To achieve those ends, we need trustworthy intermediate institutions with the right kinds of norms. The goal of regulation should be to give social media companies incentives to take on their appropriate responsibilities in the digital public sphere

NOTES

This essay was originally delivered as the keynote address of the Association for Computing Machinery Symposium on Computer Science and Law, New York City, October 28, 2019.

1 Jack M. Balkin, *Free Speech Is a Triangle*, 118 COLUM. L. REV. 2011, 2037–40 (2018).

2 See *Adult Nudity and Sexual Activity*, FACEBOOK COMMUNITY STANDARDS, https://www.facebook.com/communitystandards/adult_nudity_sexual_activity [<https://perma.cc/EFQ8-RGPW>] (last visited Mar. 10, 2020).

3 Jack M. Balkin, *Cultural Democracy and the First Amendment*, 110 NW. U. L. REV. 1053 (2016).

4 *Associated Press v. United States*, 326 U.S. 1, 20 (1945).

5 See ROBERT C. POST, *DEMOCRACY, EXPERTISE, AND ACADEMIC FREEDOM: A FIRST AMENDMENT JURISPRUDENCE FOR THE MODERN STATE* (2012) (arguing that professional and disciplinary norms for knowledge production are necessary to achieve the “democratic competence” necessary for democratic self-government).

6 SARAH T. ROBERTS, *BEHIND THE SCREEN: CONTENT MODERATION IN THE SHADOWS OF SOCIAL MEDIA* (2019); Casey Newton, *The Trauma Floor: The Secret Lives of Facebook Moderators in America*, THE VERGE (Feb. 25, 2019), <https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona> [<https://perma.cc/CBR7-DPKF>].

7 SHOSHANA ZUBOFF, *THE AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER* (2019).

8 See Balkin, *Free Speech Is a Triangle*, *supra* note 1, at 2025–27.

9 K. Sabeel Rahman, *Regulating Informational Infrastructure: Internet Platforms as the New Public Utilities*, 2 GEO. L. TECH. REV. 234 (2018).

10 See ZUBOFF, *supra* note 7, at 138–55 (describing the “Dispossession Cycle”).

11 TARLETON GILLESPIE, *CUSTODIANS OF THE INTERNET: PLATFORMS, CONTENT MODERATION, AND THE HIDDEN DECISIONS THAT SHAPE SOCIAL MEDIA* (2018).

12 See Jack M. Balkin, *Information Fiduciaries and the First Amendment*, 49 U.C. DAVIS L. REV. 1183, 1209 (2016); Jack M. Balkin, *The Three Laws of Robotics in the Age of Big Data*, 78 OHIO ST. L.J. 1217, 1228 (2018); Balkin, *Free Speech is a Triangle*, *supra* note 1; Jack M. Balkin, *The First Amendment in the Second Gilded Age*, 66 BUFFALO L. REV. 979 (2018); Jack M. Balkin, *Fixing Social Media’s Grand Bargain* (Hoover Working Group on National Security, Technology, and Law, Aegis Series Paper No. 1814, October 16, 2018), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3266942 [<https://perma.cc/SC32-HLTA>].

13 See Tarleton Gillespie, *Platforms Are Not Intermediaries*, 2 GEO. L. TECH. REV. 198, 214–16 (2018).

14 *Political Content*, TWITTER, <https://business.twitter.com/en/help/ads-policies/prohibited-content-policies/political-content.html> [<https://perma.cc/HRN2-Z4QA>] (last visited Mar. 10, 2020).

15 Rob Leathern, *Expanded Transparency and More Controls for Political Ads*, FACEBOOK NEWSROOM (Jan. 9, 2020), <https://about.fb.com/news/2020/01/political-ads/> [<https://perma.cc/S8YE-BFUL>]; Q&A on *Transparency for Electoral and Issue Ads*, FACEBOOK NEWSROOM (May 24, 2018), <https://about.fb.com/news/2018/05/q-and-a-on-ads-transparency/> [<https://perma.cc/5LWC-CUBP>].

16 See Thomas Kadri & Kate Klonick, *Facebook v. Sullivan: Public Figures and Newsworthiness in Online Speech*, 93 S. CAL. L. REV. 37 (2020).

About the Author

JACK M. BALKIN is Knight Professor of Constitutional Law and the First Amendment at Yale Law School. He is the founder and director of Yale's Information Society Project, an interdisciplinary center that studies law and new information technologies. He also directs the Abrams Institute for Freedom of Expression and the Knight Law and Media Program at Yale. He founded and edits the blog Balkinization, and has written widely on legal issues for publications including The New York Times, The American Prospect, The Atlantic, and Slate. He is the author or editor of a dozen books; most recently, he published *Democracy and Dysfunction* (University of Chicago Press, 2019) based on a series of letters between him and Sanford Levinson.

© 2020, Jack M. Balkin.

About the Knight First Amendment Institute

The Knight First Amendment Institute at Columbia University defends the freedoms of speech and the press in the digital age through strategic litigation, research, and public education. Its aim is to promote a system of free expression that is open and inclusive, that broadens and elevates public discourse, and that fosters creativity, accountability, and effective self-government.

knightcolumbia.org

Design: Point Five

Illustration: © Dale Crosby Close



**KNIGHT
FIRST AMENDMENT
INSTITUTE**

at Columbia University